SPECIFIC AIMS

Cancer is the second leading cause of death in the United States, accounting for nearly 25% of all deaths (580,000/year); in 2015, 1.7 million new cases were diagnosed. Kinase dysregulation by oncogenic mutation, amplification, or translocation plays a critical role in many of these cancers, making targeted kinase inhibition with selective inhibitors an attractive therapeutic route. Numerous difficulties frustrate the development of kinase inhibitors with cancer-targeted selectivity profiles, and there is great need for methods that speed the development of new inhibitors for novel targets and second-line therapeutics for targets that have become therapy-resistant.

Recently, the importance of protonation state effects in kinase inhibitor affinity, selectivity, and kinetics has been recognized for the most well-studied of kinase:inhibitor interactions—the binding of Abl by imatinib (Gleevec), a blockbuster drug used to treat chronic myelogenous leukemia and gastrointestinal stromal tumors. In Abl:imatinib binding, the inhibitor and kinase both populate a mixture of protonation states, and the dominant protonation states change upon binding. Changes in pH also significantly modulate binding kinetics. **Failure to account for protonation state effects can lead to significant errors of several kcal/mol in quantitative modeling of kinase:inhibitor association, needlessly frustrating selective inhibitor design efforts or investigations of the physiological efficacy of inhibitors.** Given the prevalence of pK_as near intracellular pH in kinase inhibitors and the proximity of titratable groups (especially the conserved DFG loop) in many kinases of pharmacological interest, protonation state effects may be a widespread but poorly appreciated phenomenon.

We propose a combined computational and experimental approach to assess the prevalence of protonation state effects, dissect their origin, and overcome the limitations of current modeling techniques. Using existing tools to first identify candidate kinase:inhibitor systems for study, we use both novel computational techniques and experiments capable of observing protonation state effects to examine these systems in detail. Using new algorithms to treat protonation states dynamically, we will dissect the dominant contributions to protonation state effects and eliminate current limitations in the way protonation states are handled in computational drug discovery.

AIM 1. Broadly survey kinase:inhibitor structures for evidence of significant protonation state effects. While there is reason to expect many kinase:inhibitor binding events will exhibit protonation state effects, this has only been concretely established for the well-studied Abl:imatinib system. MCCE2 (MultiConformation Continuum Electrostatics) from the Gunner lab will be used to survey kinase:inhibitor complexes for likely protonation state effects by predicting protonation/tautomer populations for all (>1480) kinase:inhibitor complexes from the PDB. This approach—which is fast, but assumes a rigid backbone is maintained—will identify complexes for subsequent detailed investigation where protonation state effects have the potential to cause large (several kcal/mol) errors in quantitative predictions of binding affinity.

AIM 2. Dissect the magnitude and nature of protonation state effects in kinase:inhibitor systems using rigorous alchemical binding free energy calculations with dynamic protonation states. We will remove the assumption that protein and ligand remain fixed in a single protonation state by rigorously incorporating dynamic sampling of protonation states into quantitatively accurate explicit-solvent alchemical free energy calculations. Inspired by efficient protonation state sampling techniques from MCCE2, we will use nonequilibrium Monte Carlo techniques capable of astronomically boosting acceptance rates to incorporate dynamic protonation state sampling into our mixed MD/MC GPU-accelerated open source free energy code. We will then computationally probe the magnitude and nature of protonation state effects for candidate systems identified in Aim 1, examining the error in computed binding affinities incurred when fixed protonation states are used for protein and/or inhibitor, as well as which species (protein or ligand, which residues/functionalities) are primary contributors.

AIM 3. Experimentally assess computational findings on kinase:inhibitor systems expected to have significant protonation state effects. We will experimentally investigate kinase:inhibitor systems predicted to have significant protonation state effects in Aim 2. We focus on bacterially-expressed human kinase domains as a well-controlled model system for assessing the accuracy of computational modeling and providing insight into the magnitude of protonation state effects on ATP-competitive inhibitor binding. We will conduct four types of experimental investigations: NMR to site-specifically identify protonation state changes in inhibitors; isothermal titration calorimetry to dissect the contribution of protonation state effects to binding affinity at physiological pH; fluorescence assays to investigate the pH-dependence of binding affinities relevant to acidified intracellular environments in cancer cells; and clinically observed mutations that may modulate binding affinities via pK_a shifts.

Outlook. This project will create the technology to directly address protonation state effects in the quantitative prediction of ligand affinities—a major roadblock to widespread application of these techniques—and provide a direct assessment of the magnitude and pervasiveness of protonation state effects in kinase-inhibitor systems.

SIGNIFICANCE

Cancer is the second leading cause of death in the United States, accounting for nearly 25% of all deaths; in 2015, 1.7 million new cases were diagnosed, and over 580,000 died [22]. In many of these cancers, kinase dysregulation is a driver and thus becomes a primary opportunity for therapeutic intervention [23]. Kinases play a critical role in the regulation of cellular signaling pathways. Perturbations of pathways due to kinase mutation, translocation, or upregulation can cause constitutive activation that drive cancer progression. Much of the effort in developing new cancer treatments (and perhaps 30% of *all* current drug development effort) is focused on designing targeted inhibitors to shut down these aberrant kinases.

Selective kinase inhibitors exert their effects with the goal of minimizing off-target interactions. These inhibitors, which specifically target one or more kinases or their mutants, have already proven their enormous potential as powerful therapeutics in the treatment of kinase dysregulation diseases. Imatinib (**Figure 1A**) specifically targets the Abl kinase dysregulated in chronic myelogenous leukemia (CML) patients to abate disease progression. Its discovery revealed the enormous therapeutic potential of selective kinase inhibitors, kindling hope that this remarkable success could be recapitulated for other cancers and diseases [24, 25].

Unfortunately, the discovery of new therapeutically useful selective molecules targeted to the kinases dysregulated in other cancers has proven more challenging. Worse yet, even when selective kinase inhibitor therapeutics are available, the inexorable emergence of resistance often limits the duration over which the patient will benefit. **Ultimately, the emergence of drug resistance where no effective therapeutics are available is thought to be the reason for treatment failure in over 90% of patients who succumb to metastatic cancer [26].** There is dire need for technologies to accelerate development of both first and second line selective kinase inhibitors.

Computational techniques hold the potential to accelerate drug discovery, but are hindered by neglect of protonation state effects. The development of imatinib (**Figure 1A**) required two years and the synthesis of over 400 molecules to refine an initial lead compound into a potent, selective inhibitor [27]. Quantitatively accurate predictive physical modeling could significantly speed the process of selective inhibitor development, especially at the lead optimization stage where potency must be improved or maintained; even a predictive accuracy of 2.0 kcal/mol could speed this process enormously, potentially cutting development time in industry from two years and $150M by up to threefold [28, 29] or permitting academic laboratories to inexpensively design selective chemical probes. Computational modeling already plays an active role in current kinase inhibitor discovery programs, but neglecting protonation state effects undoubtedly contributes to its limited effectiveness [30]. While some reports demonstrate that accounting for populations of ligand protomers enhances the performance of docking and virtual screening (e.g. [31]), a detailed understanding of the prevalence, nature, and severity of these effects is lacking, and comprehensive modeling approaches to treat them do not yet exist [32]. **This proposal aims to both establish our ability to dynamically treat protonation states and to use this as a tool to probe the prevalence, nature, and magnitude of protonation state effects in kinase:inhibitor recognition.**

Recent work has highlighted the importance of protonation state effects in imatinib:Abl binding. Aggregated experimental and computational data for the well-studied Abl:imatinib system is summarized schematically in **Figure 2**. At typical intracellular pH of 7.4, **unbound imatinib populates a mixture of protonation states**, predominantly a neutral form (0 $k_B T$) and a higher free energy protonated form where the piperazine ring is protonated ($\sim$0.6 $k_B T$) [21], as determined by electrochemical titration (**Figure 1A**). However, the complex prefers the charged form of imatinib, which incurs an unfavorable energetic cost to populate the protonated form of imatinib. With a total binding affinity of 17 $k_B T$, there is ample free energy to pay the cost of shifting the balance of protonation states from the equilibrium distribution found in solution in this complex.

(A) imatinib
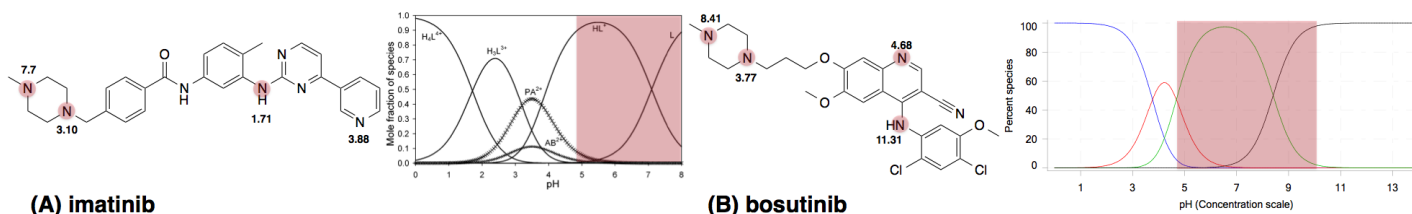
(B) bosutinib

Figure 1. **Experimentally-determined pK_as of selective kinase inhibitors imatinib and bosutinib show multiple protonation states are relevant under physiological conditions.** (A) The Abl-selective kinase inhibitor imatinib and its experimentally-determined macroscopic pK_as shown next to their corresponding titratable sites, along with microspecies populations in solution [21]. The red shaded region in the microspecies population plot shows the range of pK_a that can be accessed with a penalty of 6 $k_B T$ (3.5 kcal/mol, sufficient to shift the affinity by 400 fold). (B) The second-generation tyrosine kinsase inhibitor bosutinib (nearly equipotent to Src and Abl) and its experimentally-determined macroscopic pK_as and microscpecies populations in solution (Sirius Analytical).
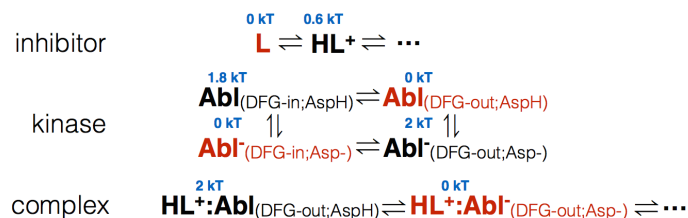
Figure 2. Scheme illustrating major protonation states of imatinib and Abl relevant to kinase:inhibitor binding at pH 7.4. An illustrative scheme depicts the two dominant protonation states of imatinib (**Figure 1A**) and protonation/conformation states of Abl, as suggested by a combination of biophysical experiments and computer simulations [21, 33, 34]. Dominant unbound and bound states are highlighted in red. Ignoring either population of a mixture of protonation states or population shifts on binding leads to errors of several $k_B T$ in computed binding free energies.

The Abl kinase also samples multiple protonation states. Asp381 in the conserved DFG (Asp-Phe-Gly) loop must swing out to permit imatinib binding. The protonation state of Asp381 is coupled to DFG loop conformation [35]. Both the cocrystal structure (pdbid:2HYY) and free energy calculations suggest the dominant bound species is the protonated form of imatinib bound to the deprotonated Asp381, while in solution it is the deprotonated imatinib and mixture of Asp conformations and protonation states [34, 36], though the exact populations of the various protonation states of Asp are still in debate [35]. A large contribution to the selectivity of imatinib for Abl over Src is actually due to the different free energy costs of reorganization of the DFG loop to permit binding, so the coupling of protonation state to conformation contributes to inhibitor selectivity as well as affinity [33, 34, 37].

Most FDA-approved kinase inhibitors can access multiple protonation states at intracellular pH. Kinase inhibitors can possess up to picomolar affinities for their targets (corresponding to a binding free energy of 28 $k_B T$, or 16.5 kcal/mol), allowing them to tolerate affinity losses of 500-fold (6 $k_B T$, equivalent to 2.6 log unit change in the equilibrium constant) while maintaining sub-nanomolar potency. As a result, kinase inhibitors can in principle access protonation states with pK_as up to 2.6 log units away from intracellular pH, allowing groups with pK_as in the range to 4.8–10 to modulate binding specificity and affinity (assuming pH$_i$ of 7.4). *More than half* of all FDA-approved kinase inhibitors have at least one predicted pK_a in this range (**Figure 3**), meaning multiple protomeric inhibitor species are likely important in the binding of numerous kinase inhibitors of clinical interest. Coupled with the potential for multiple relevant protein protonation states—possibly beyond the DFG loop alone—protonation state effects could be widespread in kinase:inhibitor recognition.

Current standard physical modeling approaches (and most drug discovery efforts) assume only a single protonation state dominates in solution and complex. In standard molecular dynamics or Monte Carlo simulations, the simulation is locked in a fixed protonation state throughout the modeling process. Even if dominant protonation states remain fixed, even a single incorrect protonation state can result in errors of several kcal/mol in computed kinase inhibitor binding affinities [34, 36, 37]. Beyond this, current techniques are incapable of treating the case where a mixture of protonation states are sampled in either solution or in in complex, and shifts in dominant protonation state upon binding (such as the case of imatinib binding) can only be correctly accounted for if the cost of populating these protonation states can be accurately assessed. As a result **any departure from a single dominant protonation state for both kinase and inhibitor in both complex and solution—or a misassignment of protonation states—will make the neglect of protonation state effects a dominant contribution to the error in predictive physical modeling and design using current techniques.**

While there have been some efforts to account for protonation state effects (such as the use of protonation state penalties in rescoring of docked poses [38], the development of constant-pH molecular dynamics in implicit [39, 40] and explicit [41–44], and even targeted enveloping distribution sampling of single protonation states in a free energy calculation [45]), **no current approaches permit fully dynamic protonation states of all relevant titratable groups in the computation of ligand-binding affinities**.

The objective of this proposal is to overcome current limitations in physical modeling approaches by integrating a dynamic treatment of protonation states into quantitatively accurate alchemical binding free energy calculations, which are currently the gold standard approach to quantitatively accurate physical modeling [9, 46]. We adopt a combined computational-experimental approach to validate this approach against quantitative experimental data. Next, we utilize this tool alongside experimental biophysical techniques to address the following questions:

1. How pervasive are protonation state effects in kinase:inhibitor affinity and selectivity?
2. What is the biophysical nature of protonation state effects in kinase inhibitor recognition? Are mixtures of protonation states populated, or shifts in dominant protonation state upon binding? Does the protein or ligand predominantly experience protonation state effects, or are they coupled?
3. What is the magnitude of the error incurred if protonation state effects are neglected?

Scientific premise: Despite uncertainty over the magnitude of these effects in Abl:imatinib—the only kinase:inhibitor system studied in detail so far—the aggregate of multiple experimental and computational studies suggests the scientific premise for the existence of protonation state effects is strong. In this project, we aim to establish the prevalence, magnitude, and nature of these effects much more broadly.

INNOVATION

Our proposed survey of kinase:inhibitor interactions for protonation state effects will provide the first estimate of how prevalent protonation state effects are in selective kinase inhibitor recognition. While detailed biophysical and computational studies of protonation state effects have been undertaken for a few specific protein model systems by our collaborator Paul Czodrowski (see included Letter of Support) in his work with Klebe [47–50], little is known about the prevalence of protonation state effects in inhibitor binding in general [32], and essentially nothing is known of the prevalence, nature, and magnitude of these effects in kinase inhibitor recognition outside of the well-studied Abl:imatinib system. The survey and subsequent detailed study we propose will greatly advance our understanding of the prevalence of protonation state effects.

We aim to implement a dynamic, physically accurate treatment of both protein and small molecule protonation states within a rigorous statistical mechanical framework, removing a significant barrier to accurate affinity predictions. While a number of methodologies have been developed for constant-pH protein simulations [11,13,14,39,41,41–44, 51,52], **this project will be the first time a dynamic treatment of all protein and small molecule protonation states has been applied to rigorous calculations of kinase inhibitor binding affinity and selectivity.**

We will use the rigorous framework of modern alchemical free energy calculations. Alchemical methods are a rigorous approach to computing protein-small molecule binding affinities at moderate computational cost [5, 9,53–56]. They employ molecular simulations—molecular dynamics (MD) or Monte Carlo (MC)—of unphysical, *alchemically perturbed* intermediate states that attenuate the interactions of the small molecule with its environment. These alchemically perturbed states include both the fully-interacting ligand-protein complex as well as replicas in which the ligand does not interact with its receptor, and allow the total free energy of binding—including entropic and enthalpic contributions—to be efficiently computed [9,55,56]. Currently, alchemical methods can achieve semi-quantitative, predictive accuracy ($\sim$1–2 kcal/mol) for well-behaved proteins in which there are no protonation state effects [5,9,46], which is sufficient accuracy to greatly accelerate lead optimization [28]. While alchemical methods are somewhat demanding, requiring several wall clock hours/ligand rather than seconds/ligand for docking, the results are both more accurate and provide more physical information.

We are uniquely positioned to implement and validate dynamic protonation state free energy calculations. This project combines the extensive expertise from the Gunner lab [11–16] in treating protein protonation states and their coupling with conformation and ligand binding with the expertise of the Chodera lab [1–10] in rigorous, physically accurate alchemical free energy methodologies. The Gunner lab has developed a highly efficient Monte Carlo code that samples protonation and tautomeric states subject to a rigid backbone approximation [11–16]. The Chodera lab has developed a GPU-accelerated, Python-based, open source alchemical free energy calculation code YANK [57]—the first GPU-accelerated alchemical free energy code [10]—designed with flexibility in mind, allowing facile implementation of new algorithms. In addition to using the GPU-accelerated OpenMM library to achieve high speed on consumer-grade graphics processors (which can achieve over 200 ns/day on the JAC DHFR PME benchmark) [58,59], this code incorporates many modern methdologies, such as Hamiltonian replica exchange with Gibbs sampling to increase phase space mixing [8], state-of-the-art reweighting techniques to extract all information from all thermodynamic states [6], the inclusion of anisotropic long-range dispersion corrections [3], and mixes Monte Carlo and molecular dynamics for efficient sampling of ligand poses [10]. In addition, the Chodera lab has developed highly efficient techniques based on nonequilibrium proposals to incorporate Monte Carlo moves into molecular dynamics simulations and increase acceptance rates astronomically [7]—a technology required to enable protein protonation state sampling in explicit solvent [44]. Acceptance rates can be further increased by optimizing alchemical nonequilibrium switching protocols using estimators we developed for this purpose [60], as well as efficient reweighting techniques capable of making use of all simulation data [6,61].

This proposal uses a robust combination of biophysical techniques to validate and complement computational approaches to assessing protonation state effects. The Seeliger lab brings extensive expertise in kinase expression, biochemistry, and structural biology—especially the quantitative NMR techniques that will be utilized in this project. Seeliger developed a widely shared and cited bacterial expression system that allows rapid generation of large quantities of pure, active kinase for biophysical studies [62]. The Seeliger lab was first to demonstrate a significant role for protonation state effects in Abl:imatinib association [35]. The Chodera lab performs automated biophysical experiments using a robotic wetlab capable of engineering many point mutations and assessing binding affinities via automated fluorescence or ITC experiments. Collaborator Paul Czodrowski contributes his extensive experience in essentially all other combined experimental/computational studies of the role of protonation state effects in small molecule recognition [47–50], as well as his expertise on current practices with regard to protonation states in active drug discovery projects. The Gunner lab also brings considerable knowledge of experimental ligand binding affinity measurements [16–18].

APPROACH

To address the role of protonation state effects—defined as the presence of a mixture of protonation states or a shift in dominant protonation states upon binding—in kinase:inhibitor association, we have two primary goals:

1. Develop and validate a rigorous computational framework to incorporate the effects of protein and small molecule dynamic protonation states within the context of alchemical free energy calculations to compute accurate protein:ligand binding affinities
2. Utilize computational and experimental biophysical approaches—including this new framework—to assess the pervasiveness, nature, and magnitude of protonation state effects in kinase inhibitor binding

To achieve these goals, the work is organized into three main Aims that successively filter kinase:inhibitor complexes into successively refined subsets that permit their detailed study. For tractability, we focus on kinase domains—the therapeutic target of kinase inhibitors—in isolation, rather than full-length kinases.
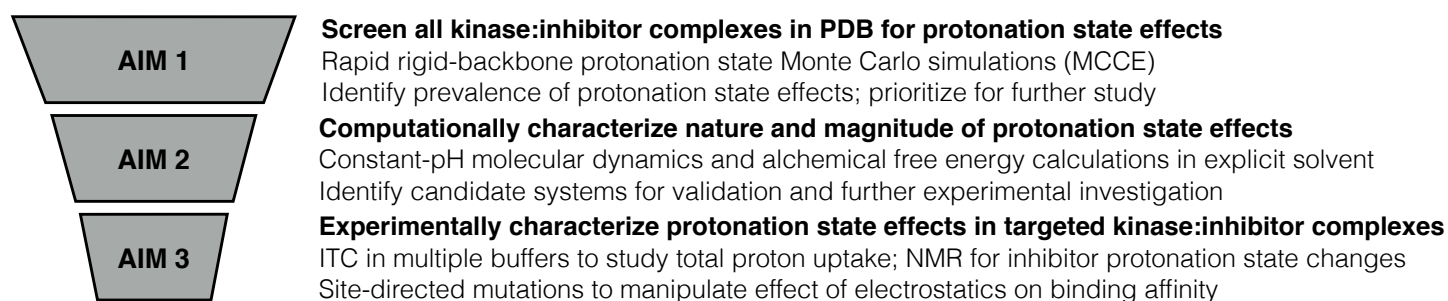
In **Aim 1**, we first construct an automated pipeline for an existing tool—MCCE2 [12, 14] from the Gunner lab—in a manner that allows us to screen *all* kinase:inhibitor complexes in the PDB to identify those where protonation state effects may play a large role in binding, identifying a subset of these (∼100–200) for further study in **Aim 2**. While this approach uses a fixed backbone and a finite library of rotamers, these approximations are required in order to identify kinase:inhibitor pairs beyond the well-studied Abl:imatinib system that might exhibit significant protonation state effects. This Aim requires an initial benchmark of the accuracy of pK_a prediction methodologies, which will utilize experimental pK_a data for all FDA-approved kinase inhibitors, informing future modeling studies that utilize predicted pK_as of novel compounds. This work will also provide the first large-scale survey of how widespread protonation state effects might be in kinase:inhibitor association.

In **Aim 2**, we incorporate dynamic protein and ligand protonation states into GPU-accelerated alchemical free energy calculations. This approach will be applied to the kinase:inhibitor systems identified as having the potential for significant protonation state effects in **Aim 1**—along with a number of controls which are believed to not display protonation state effects. Kinase:inhibitor systems that are predicted through this much more detailed atomistic approach to have experimentally measurable effects (significant ligand protonation state population shifts for NMR, significant net proton uptake/loss events for the kinase:inhibitor complex for ITC; and pH- or mutation-sensitive binding affinity changes for fluorescence assays) will be utilized in **Aim 3** for validation. In parallel, dynamic protonation state free energy calculations will be used to validate the prevalence of protonation state effects (by confirming the number of kinase:inhibitor systems that exhibit them), the nature of these effects (by identifying whether protein, inhibitor, or both exhibit the effects), and magnitude of these effects (by addressing the magnitude of error made if protonation state effects are neglected for protein, inhibitor, or both).

In **Aim 3**, we use multiple experimental biophysical techniques to both validate the accuracy of the dynamic protonation state alchemical free energy methodology developed in **Aim 2** as well as confirm its findings regarding the prevalence, nature, and magnitude of protonation state effects. We focus on experimentally tractable systems, favoring kinases that can be easily expressed and inhibitors that can be easily obtained. NMR techniques will be used to examine cases where the protonation state of the ligand is expected to undergo a detectable protonation state change upon binding. ITC techniques that utilize multiple buffers at the same pH will be used to dissect the contribution protonation state changes to binding in cases where net proton uptake/loss is expected and inhibitor is sufficiently soluble. Fluorescence assays to measure direct inhibitor binding affinities will be used in cases where the affinity is predicted to be sensitive to pH or engineered point mutations.

Relevant biological variables have been considered in the design of this study, including posttranslational modifications, choice of protein constructs, concentrations and buffer conditions, and the effects of pH.

Scientific rigor is ensured by use of established experimental protocols, well-characterized kinase inhibitors, and validation of computational, structural, and experimental data through appropriate statistical methods.

AIM 1 — **Screen all kinase:inhibitor complexes in PDB for protonation state effects**
Rapid rigid-backbone protonation state Monte Carlo simulations (MCCE)
Identify prevalence of protonation state effects; prioritize for further study

AIM 2 — **Computationally characterize nature and magnitude of protonation state effects**
Constant-pH molecular dynamics and alchemical free energy calculations in explicit solvent
Identify candidate systems for validation and further experimental investigation

AIM 3 — **Experimentally characterize protonation state effects in targeted kinase:inhibitor complexes**
ITC in multiple buffers to study total proton uptake; NMR for inhibitor protonation state changes
Site-directed mutations to manipulate effect of electrostatics on binding affinity

AIM 1. Broadly survey kinase:inhibitor structures for evidence of significant protonation state effects.
While there is reason to expect many kinase:inhibitor binding events will exhibit protonation state effects, this has only been concretely established for the well-studied Abl:imatinib system. MCCE2 (MultiConformation Continuum Electrostatics) from the Gunner lab will be used to survey kinase:inhibitor complexes for likely protonation state effects by predicting protonation/tautomer populations for all (>1480) kinase:inhibitor complexes from the PDB. This approach—which is fast, but assumes a rigid backbone is maintained—will identify complexes for subsequent detailed investigation where protonation state effects have the potential to cause large (several kcal/mol) errors in quantitative predictions of binding affinity.

Rationale: Aside from the well-studied Abl:imatinib system, little is known about the degree to which protonation state effects are relevant in kinase:inhibitor recognition. These effects could include both changes in the population mixtures of protomeric species and shifts in dominant protomeric species upon binding, either of which can cause large (several kcal/mol) errors in modeling. Using multiconformer continuum electrostatics methods in MCCE2 [12, 14] developed by the Gunner group for the study of protonation state effects, we will survey kinase:inhibitor complexes from the PDB for evidence of significant protonation state effects in ligand or protein. Systems with significant protonation state population changes will be targeted in subsequent Aims.

Aim 1A. Benchmark small molecule pK_a calculation methodologies against experimental kinase inhibitor data

Approach: To survey the >1480 kinase:inhibitor complexes in the PDB for potential protonation state effects, we need to predict potential inhibitor protonation states and their relative populations in solvent. To select an appropriate predictive scheme, we will first benchmark the accuracy of a number of pK_a methodologies readily available to us as stand-alone software packages, such as Epik [63], Jaguar (Schrödinger), MoKa [64], ChemAxon Physico-chemical property predictors (ChemAxon), and ACD/pKa (ACD/Labs). For this benchmark, we will utilize 22 FDA-approved small molecule kinase inhibitors, which are heavily represented in kinase:inhibitor structures in the PDB. When experimental data is unavailable, we will collect new pK_as data using electrochemical and UV-metric acid-base titrations using a Sirius T3 (via Sirius Analytical), with ambiguous site specificity reconciled by ^{1}H-NMR titrations (using the same methodology in **Aim 3A**).

Preliminary results: We used the Epik [63] tool to predict pK_as for 22 FDA-approved kinase inhibitors (**Figure 3**), with pK_as of imatinib and bosutinib reproduced to <1 log unit error. More than half of these inhibitors are predicted to possess pK_as in a range that would require a free energy cost of 6 k_BT or less to access an alternative protonation state at the typical intracellular pH of 7.4, suggesting protonation state effects may be prevalent among selective kinase inhibitors.
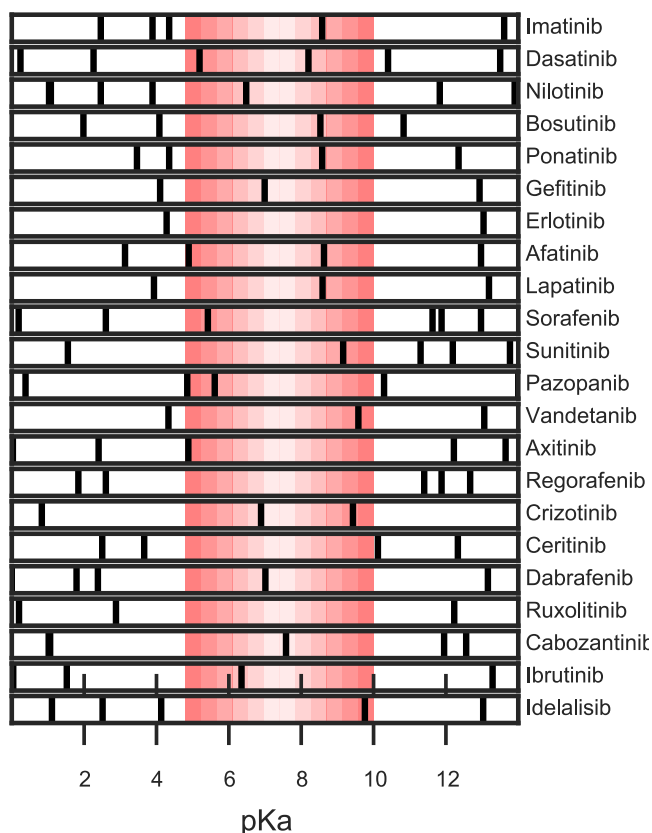
Figure 3. **More than half of all FDA-approved small molecule kinase inhibitors have multiple relevant protonation states.** pK_as predicted by Epik [63] (Schrödinger) are shown. The red region denotes the pK_a range most likely to cause protonation state effects, with each progressively darker region costing an additional k$_B$T to populate at an intracellular pH of 7.4, up to maximum of 6 k$_B$T. Kinase inhibitors can possess up to picomolar affinities for their targets (corresponding to a binding free energy of 28 k_BT, or 16.5 kcal/mol), allowing them to tolerate affinity losses of 500-fold (6 k_BT, equivalent to 2.6 log unit change in the equilibrium constant) while maintaining sub-nanomolar potency. As a result, kinase inhibitors can in principle access protonation states with pK_as up to 2.6 log units away from intracellular pH, allowing groups with pK_as in the range to 4.8–10 to modulate binding specificity and affinity (assuming pH$_i$ of 7.4).

Aim 1B. Survey noncovalent kinase:inhibitor complexes for significant protonation state effects using MCCE2.

Approach: We will extract all noncovalent kinase:inhibitor complexes in the PDB to perform a large-scale survey of potential protonation-state effects using MCCE2 [12, 14]. MCCE2 uses Monte Carlo (MC) sampling of sidechain rotamers and sidechain/inhibitor protonation states to efficiently sample the equilibrium distribution of protomeric states [12, 14]. While MCCE2 makes the approximation that the protein backbone and ligand geometry is fixed for efficient sampling, protonation states, tautomers, and sidechain rotamers are sampled during the simulation.

All kinase inhibitors will have solution protonation and tautomer populations predicted as in Aim 1A and charges assigned using the AM1-BCC method [65, 66] to produce charges compatible with the MCCE2 titratable protein forcefield. For each complex, a complete model of the kinase catalytic domain (with missing heavy atoms added and internal loops modeled) will be generated using MODELLER [67, 68], and an explicit-hydrogen form of the ligand constructed using the OpenEye toolkit. MCCE2 will be run on inhibitor alone, kinase alone, and complex to identify (1) whether mixtures of protonation states exist in significant populations in any of these states, and (2) if so, whether dominant protonation states of inhibitor or kinase are seen to shift upon binding. MCCE requires ∼6-12 hr/complex on a single processor, allowing all ∼1400 complexes to be rapidly evaluated. Kinase:inhibitor pairs that appear to have large protonation state effects despite these assumptions will be flagged as candidates for subsequent detailed study in **Aim 2**.

Preliminary data: We performed a cursory survey of 10 kinase:inhibitor complexes in the RCSB using MCCE2 in 'quick' mode, where limited conformational sampling of sidechain rotamers is performed. Protonation state penalties and charges were obtained as described above. **Table 1** summarizes the results of this survey. Protonation state effects leading to large modeling errors can arise from either a shift in dominant protonation state upon binding or from the population of a mixture of protonation state in either complex or solution. Of the 10 complexes surveyed, only 1/10 kinase:inhibitor complexes did not show any predicted shift in average protein or inhibitor protonation state upon binding; in 1/10 cases, the kinase experienced a large (+0.9) shift, while in 4/10 cases, the inhibitor experienced a large (>0.5) shift in Δn_H. For the inhibitor, only 2/10 cases did not populate a mixture of protonation states or experience a large shift in dominant protonation state upon binding.

| | | | Δn_H on binding | | inhibitor protonation state populations | |
kinase	inhibitor	PDBID	kinase	inhibitor	in solution	in complex
EGFR	afatinib	4G5J	-0.35	**-0.83**	**98.0% / —**	**19.0% / 80.5%**
ALK	alectinib	3AOX	-0.14	+0.09	**64.8% / 35.2%**	**56.1% 43.9%**
VEGFR2	axitinib	4AG8	+0.00	+0.00	**82.7% / 17.4%**	**— / 100%**
ABL1	bosutinib	3UE4	-0.04	+0.00	88.7% / — / — / —	97.5% / — / — / —
ALK	ceritinib	4MKC	**+0.90**	+0.00	100%	100%
MEK1	cobimetinib	4AN2	-0.30	+0.00	**— / 95.8%**	**88.8% / —**
MET	crizotinib	2WGJ	-0.17	**-0.80**	**62.0% / 37.3% / — / —**	**— / — / 76.4% / 23.6%**
ALK	crizotinib	4ANQ	-0.15	**-0.60**	**62.0% / 37.3% / —**	**— / — / 99.5%**
ABL1	ponatinib	3IK3	-0.28	**+0.50**	**57.6% / 38.4%**	100% / —
DDR1	ponatinib	3ZOS	-0.35	+0.00	**55.2% / — / 41.3%**	**— / 97.4% / —**

Table 1. **Preliminary MCCE2 calculations of net proton gain or loss (Δn_H) upon inhibitor binding at pH 7.4 for kinase and inhibitor for selected kinases.** Ten kinase:inhibitor complexes from the RCSB were surveyed using Epik [63] protonation state energy penalties for the inhibitor at pH 7.4 and AM1-BCC charges [65, 66] using the 'quick' mode, where sampling of sidechain rotamers is not fully exhaustive. Of these, only VEGFR2:axitinib did not show any predicted shift in average protein or inhibitor protonation state upon binding. Most other complexes exhibited mild ($|\Delta n_H| < 0.5$) protonation state shifts in protein (8/10) or inhibitor (1/10), while a few showed large ($|\Delta n_H| \geq 0.5$) shifts in average protonation state of kinase (1/10) or inhibitor (4/10) as shown in **bold**. Initibitor protonation states in solution and in complex are also shown, with the population of each protonation state listed; small ($< 5\%$) populations are denoted with a dash —. Significant protonation state effects for the inhibitor occur when the inhibitor populates a mixture of protonation state in either solution or complex (7/10), or if there is a shift in dominant protonation state (6/10); these cases are denoted in **bold**. Statistical uncertainty in populations was ∼0.1%, but the limited rotamer variety used in these pilot calculations introduces additional error.

Expected outcomes: The survey described in this Aim will provide the first estimate of the overall importance of protonation effects in kinase:inhibitor binding, and will provide numerous candidates for future in-depth study. It will also provide a critical benchmark of quantitative accuracy for pK_a prediction methods for kinase inhibitors.

Potential pitfalls and alternative approaches: Crystal structures of Abl:imatinib will be used as a control, since protonation state effects have been characterized for this system. Should pK_a predictions for kinase inhibitors prove too inaccurate, we will focus subsequent Aims on kinase inhibitors with available experimental pK_a data. If too many potential kinase:inhibitor systems may have significant protonation state effects, we will prioritize kinases with facile bacterial expression and kinase inhibitors that can be easilyobtained to facilitate subsequent experimental Aims. Multiple simulation runs and statistical error analysis will ensure reproducibility of calculations.

AIM 2. Dissect the magnitude and nature of protonation state effects in kinase:inhibitor systems using rigorous alchemical binding free energy calculations with dynamic protonation states. We will remove the assumption that protein and ligand remain fixed in a single protonation state by rigorously incorporating dynamic sampling of protonation states into quantitatively accurate explicit-solvent alchemical free energy calculations. Inspired by efficient protonation state sampling techniques from MCCE2, we will use nonequilibrium Monte Carlo techniques capable of astronomically boosting acceptance rates to incorporate dynamic protonation state sampling into our mixed MD/MC GPU-accelerated open source free energy code. We will then computationally probe the magnitude and nature of protonation state effects for candidate systems identified in Aim 1, examining the error in computed binding affinities incurred when fixed protonation states are used for protein and/or inhibitor, as well as which species (protein or ligand, which residues/functionalities) are primary contributors.

Rationale: Current modeling techniques for protein-ligand interactions assume both protein and ligand retain a single, dominant protonation state—an assumption likely violated by many selective kinase inhibitors of clinical interest. We will remove this limitation by extending alchemical free energy calculations—which allow the free energy of inhibitor binding to be rigorously computed—to permit dynamic sampling of protonation states. Integrating efficient MC techniques for sampling protonation states (from the Gunner lab's MCCE2 [12,14]) into the mixed MD/MC alchemical binding affinity computation code YANK [10] from the Chodera lab using nonequilibrium Monte Carlo techniques [7,44], we can account for the thermodynamic effects of protein and small molecule dynamic protonation states in explicit solvent in a quantitative, rigorous manner. We will then computationally probe the magnitude and nature of protonation state effects for candidate systems identified in **Aim 1**.

Aim 2A. Implement constant-pH sampling capabilities into the YANK *alchemical free energy calculation code.*

Approach: The Chodera laboratory has developed a free, open-source, GPU-accelerated code YANK [57] for computing protein-ligand binding affinities using rigorous alchemical free energy methods. YANK combines numerous efficiency and accuracy improvements previously developed by Chodera and collaborators [1–4,6–8,69] into a single code with an open architecture that allows new algorithms to be easily developed, implemented, and validated. YANK has recently been shown to accurately recover both geometrical binding orientation and numerical ligand binding affinities for a protein system where there is only a single relevant protonation state [10].

We will implement a dynamic, rigorous, consistent treatment of both protein and small molecule protonation states into our alchemical free energy code. YANK uses a Gibbs sampling replica exchange molecular dynamics (MD) framework, where many exchange attempts among alchemical intermediates are periodically attempted and accepted with a Metropolis-like criterion to give correct equilibrium statistics [8,10]. The MCCE2 program from the Gunner lab uses Monte Carlo (MC) sampling of sidechain rotamers and protonation states to efficiently sample the appropriate equilibrium distribution of protomeric states [12,14]. As YANK already mixes MC with MD for enhanced sampling [10]), it is straightforward to incorporate MCCE2-like protonation state sampling into YANK to introduce stochastic discrete protonation (and proton tautomer) state changes. While there is precedent for hybrid MD/MC protonation state sampling in implicit [11,13,14,39] and explicit [41,44] solvent to account for electrostatic interactions that influence protomer populations, this would be the first instance of this technique utilized in alchemical binding free energy calculations. We will use our highly efficient NCMC approach [7] to enable efficient sampling of protonation states of multiple titratable groups in both proteins and small molecules using either implicit solvent (which trades some accuracy for speed) and explicit solvent (the current gold standard), and use theromodynamic length techniques top optimize protonation state switching protocols [60,70,71].

During the simulation, MC moves attempt to switch the current protonation (or proton tautomer) state of a residue or cluster of residues to another discrete, physical protonation state. Proposals can be biased by inexpensive MCCE2-like MC moves to increase acceptance rates. MC moves will utilize NCMC (nonequilibrium candidate Monte Carlo) [7], in which the initial configuration is stored, a short velocity Verlet [72] trajectory of n steps is generated in which the protonation state incrementally changes to its new proposed discrete state through some (possibly alchemical) nonequilibrium protocol in which the rest of the system is also allowed to respond to the protonation state change. The final configuration and physical, discrete protonation state is either accepted according to a Metropolis-like criteria involving the current potential energy, the solvent pH, and the solution population of the protomeric state. If rejected, the entire switching trajectory is rejected, allowing rigorous sampling the correct constant-pH ensemble—the intermediate nonphysical protonation states are never recorded [7]. In explicit solvent, a water is reversibly turned into a monovalent counterion to preserve charge neutrality.

The key to alchemical free energy calculations with dynamic protonation states is defining an appropriate *reduced potential* $u(x, \mathbf{s}; \lambda, \mathrm{pH})$, where the equilibrium distribution sampled is $\pi(x, \mathbf{s}; \beta, p, \lambda, \mathrm{pH}) \propto e^{-u(x,\mathbf{s};\beta,p,\lambda,\mathrm{pH})}$, with x the current instantaneous configuration, $\mathbf{s} \in \mathbb{Z}^N$ an integer vector specifying the current protonation state of all N titratable sites for inhibitor and protein, $\beta = 1/(k_B T)$ the inverse temperature, p pressure, $\lambda \in [0, 1]$ the alchemical parameter, and pH denoting the current pH. We use the reduced potential

$$u(x, \mathbf{s}; \beta, p, \lambda, \mathrm{pH}) \equiv \beta \left[U(x; \lambda) + pV(x) + \sum_{n=1}^{N} g_{n,s_n}(\mathrm{pH}) \right] \tag{1}$$

where $U(x; \lambda)$ is the potential energy function, $V(x)$ the current volume, and $g_{n,s}(pH)$ are dimensionless log weights calibrated to ensure that the corresponding reference compound for the titratable site gives the appropriate equilibrium distribution of protonation states in solvent at the pH of interest. The equilibrium distribution can be sampled by hybrid Monte Carlo/molecular dynamics methods in an expanded ensemble context [7, 8, 73] within a Gibbs-sampling Hamiltonian replica exchange simulation for the alchemical λ [8]. For amino acids, the reference compound is the terminally-blocked form of the amino acid, and experimental pK_as are used to calculate the target populations at the pH of interest. For small molecules, equilibrium solvent protonation state populations at the pH of interest are obtained from pK_a prediction software assessed in **Aim 1A**. Calibration of the log weights $g_{n,s}(\mathrm{pH})$ uses self-adjusted mixture sampling (SAMS) [74], and must be repeated for each combination of pH, temperature, pressure, solvent/forcefield/charge model, and nonbonded electrostatics method. These log weights incorporate both a pH-dependent chemical potential difference between protons, waters, and counterions, as well as the molecular mechanics valence energy differences between protomer/tautomer species. Because small molecules may have multiple protonation sites coupled through quantum mechanical effects that cannot be captured in the purely electrostatic model of site-site coupling used here to shift protomer populations, the entire small molecule inhibitor is treated as a single titratable site with many possible protomers with different partial charges.

The free energy contribution for each alchemical leg is optimally estimated using the multistate Bennett acceptance ratio (MBAR) method we developed to analyze equilibrium data sampled from arbitrary equilibrium thermodynamic states [6], by solving the self-consistent equations for the dimensionless free energies $\hat{f}_i$, $i = 1, \ldots, K$,

$$\hat{f}_i \equiv -\ln \sum_{k=1}^{K} \sum_{n=1}^{N_k} \left[\sum_{l=1}^{K} \exp\left\{ \hat{f}_l - [u(x_{kn}, \mathbf{s}_{kn}; \beta_l, p_l, \lambda_l, \mathrm{pH}_l) - u(x_{kn}, \mathbf{s}_{kn}; \beta_i, p_i, \lambda_i, \mathrm{pH}_i)] \right\} \right]^{-1} \tag{2}$$

where there are N_k samples x_{kn} from each of K alchemical states. It is important to note that this approach differs substantially from both the enveloping distribution sampling approach to alchemical free energy calculations with post-simulation protonation state evaluation [45] which is unable to scale to treat many dynamic protonation states for both protein and ligand, and differs from existing approaches to constant-pH molecular dynamics in explicit solvent [41–44] which have not yet been rigorously incorporated into alchemical free energy calculations.

Potential pitfalls and alternative approaches: No significant impediments are expected due to the straightforward nature of this Aim. There is some uncertainty surrounding the NCMC efficiency; based on earlier work [41, 44], we expect nonequilibrium switching times of 10–50 ps to be sufficient without optimization of the switching pathway. Given the efficiency of the the GPU-accelerated OpenMM library [58] on which YANK is based, simulations achieve 200–300 ns/day for DHFR in explicit solvent, which is expected to allow for 10^4–10^5 protonation state trials/day in a standard 10-GPU alchemical free energy calculation. Should convergence of dynamic protonation states prove problematic, optimization of the switching pathway and incorporation of MCCE-like proposal biasing will further increase efficiency. Control experiments—such as the computation of equilibrium protomer populations and pK_as for blocked amino acids and small molecules—will ensure the correctness of the implementation.

Aim 2B. Assess the prevalence, magnitude, and nature of protonation state effects in kinase inhibitor binding

A central aim of this proposal is to determine the extent to which protonation state effects play a role in the recognition of selective kinase inhibitors. We will apply the dynamic protonation state free energy code developed in **Aim 2A** to the kinase:inhibitor complexes identified in **Aim 1B** expected to exhibit protonation state effects in order to assess the **prevalence** (fraction of kinase:inhibitor complexes that display significant protonation state effects), **nature** (which residues or chemical moieties are primarily responsible for these effects, and to what degree there is coupling between titratible groups in kinase and inhibitor), and **magnitude** (error induced by neglect of dynamic protonation states in kinase, inhibitor, or both) of protonation state effects in these systems.

We will focus on complexes identified as likely candidates from **Aim 1B**—along with controls (Abl:imatinib and several complexes identified to be absent protonation state effects)—with an estimated 100–200 total complexes considered. (100 complexes $\times$ 10 GPUs/complex $\times$ 10 days/complex = 5000 GPU-days on the MSK 120-GPU cluster, or a total of 82 cluster-days [or fewer if GPUs are upgraded] over the course of this project.) We will primarily consider the dephosphorylated form of protein complexes, though we will examine the effects of phosphorylated forms of some complexes for which the experimental phosphorylation state is known. If suitable expression constructs are available (for **Aim 3**), simulations will utilize the same protein construct available for expression. When multiple experimental constructs are available, sensitivity to construct choice will be assessed.

We will quantify the impact of protonation state effects on inhibitor binding affinities and selectivities.
Many kinase:inhibitor binding affinities have been reported in the literature (e.g. [75–77]), measured in the Chodera or Seeliger laboratories, or will be measured as part of Aim 3. In addition, a number of kinase:inhibitor complexes available in the PDB represent the same inhibitor complexed with different kinases—for example, structures of imatinib with both Src and Abl are available, despite the high selectivity for Abl over Src. For cases where experimental affinity data is available, we will assess the degree to which the inclusion of protonation state effects increases the accuracy in reproducing experimental inhibitor binding affinities. We will ask whether small molecule protonation state effects are crucial in discriminating a selective binder from a non-selective binder. We will assess whether neglect of accurate or dynamic protonation states in the protein and/or ligand leads to significant shifts in affinity, and assess their magnitude. We expect better understanding of when and where protonation effects influence drug binding will aid active drug discovery projects and inform the development of future accurate methodologies to guide both early-stage lead discovery and late-stage lead optimization efforts.

Potential pitfalls and alternative approaches: Slow conformational changes—especially those coupled to protonation state changes—can impede convergence in constant-pH simulations. To mitigate this risk, we can perform calculations starting with modeled active and inactive conformations for tyrosine kinases (following [34]) using our Ensembler automated modeling tool [78] to assess dependence of calculated affinities and protonation states on conformation. Notably, the accuracy of these calculations will depend on the protein and small molecule forcefields. We will benchmark multiple forcefields, starting with the AMBER family of forcefields [79]. If forcefield inaccuracies prove problematic at replicating experimental affinities, our calculations should provide qualitative insight into the impact of protonation state effects. Even if few significant protonation state effects are identified, this knowledge will aid in identifying other dominant sources of limiting accuracy in modeling tools.

> **AIM 3. Experimentally assess computational findings on kinase:inhibitor systems expected to have significant protonation state effects.** We will experimentally investigate kinase:inhibitor systems predicted to have significant protonation state effects in Aim 2. We focus on bacterially-expressed human kinase domains as a well-controlled model system for assessing the accuracy of computational modeling and providing insight into the magnitude of protonation state effects on ATP-competitive inhibitor binding. We will conduct four types of experimental investigations: NMR to site-specifically identify protonation state changes in inhibitors; isothermal titration calorimetry (ITC) to dissect the contribution of protonation state effects to binding affinity at physiological pH; fluorescence assays to investigate the pH-dependence of binding affinities relevant to acidified intracellular environments in cancer cells; and clinically observed mutations that may modulate binding affinities via pK_a shifts.

Rationale: In this Aim, we will experimentally investigate those kinase:inhibitor systems predicted to have significant protonation state effects in **Aim 2**, using our analysis of the results of **Aim 2B** to identify systems in which different experimental measurements would be most appropriate for detecting observable protonation state effects. For experimental tractability, we restrict ourselves to kinases for which bacterial expression permits both large-scale production and the introduction of site-directed mutants, and inhibitors for which compounds can readily be obtained or purchased from vendors. The Chodera and Seeliger labs have applied the phosphatase coexpression strategy of pioneered by Seeliger [62] and demonstrated good bacterial expression ($\geq$ 2 mg/L culture) for 52 human kinase catalytic domains with available crystal structures using an automated expression pipeline [80], providing ample opportunities for kinases to be carried forward for experimental investigation. We will primarily consider the dephosphorylated form of these kinases, though we will examine the effects of phosphorylation for some kinases for which the biochemically-accessible phosphorylation state is known. Notably, we need not wait to complete **Aim 2B** before promoting promising complexes for study in this experimental stage—these experiments can be run in parallel once candidate kinase:inhibitor complexes are identified.

Aim 3A. Examine changes in inhibitor protonation state upon binding using NMR

^{1}H-NMR is a powerful technique for directly observing protons in small molecules. Indeed, this technique has been used to identify site-specific pK_as for imatinib and its chemical fragments to complement potentiometric titration [21]. By titrating the inhibitor or inhibitor:kinase complex over a pH range, the pH-dependent peak integrals can be compared with computational predictions of average fractional protonation, or the effective microscopic pK_as fit. The Seeliger lab will make use of a high-field NMR instrument with cryoprobe (3-5× more sensitive than the instrument used in [21]). To suppress the ^{1}H signal of the protein when measurements of the complex are made, we will use saturation transfer difference (STD) NMR [81]. Comparison of measured and computed pH titration curves and pK_as will assess the accuracy of the alchemical free energy calculations, as well as provide direct experimental evidence for the presence or absence of inhibitor protonation state effects.

Selection criteria: Kinase:inhibitor complexes identified in **Aim 2B** to experience significant shifts in inhibitor protonation state upon binding will first be subjected to a pH-dependent alchemical free energy calculation, in which a 2D grid of thermodynamic states in which alchemical λ parameter and pH are both varied and analyzed collectively with MBAR to determine pH-dependent inhibitor protonation states (similar to [51]). Inhibitors predicted to show a measurable change in protonation state in complex over the accessible pH range (for which the kinase is stable) will be identified as candidates for these experiments.

Potential pitfalls and alternative approaches: Kinases are often stable only over a limited pH range, with pH < 6 becoming especially difficult. To ascertain the range of stability, will perform Thermofluor melts over a range of buffer pH using the Chodera lab automation system [82], as this technique consumes little protein but can give accurate readouts of stability via melting temperatures. Should STD-NMR pH titrations prove unworkable for suppressing protein ^{1}H signal while still observing ligand protonation signals, we will fall back to complete kinase deuteration, though the expense of this will limit the number of kinases we can investigate via NMR.

Aim 3B. Observe protonation uptake/loss using ITC, and dissect its contribution to affinity

Isothermal titration calorimetry (ITC) provides a direct route to studying proton shuttling between solvent and complex upon binding. ITC measurements in buffers with identical pH and ionic strength but distinct ionization enthalpies (e.g. Tris, Hepes, and Tricine), allow the heat effects of proton shuttling between complex and solvent to be deconvoluted from heat effects of binding, permitting experimental observation of protonation state effects [83]. While experiments of this kind have not yet been conducted on kinase inhibitor binding, co-I Seeliger has previously demonstrated standard ITC experiments are possible several kinase:inhibitor systems [33]. These experiments will be conducted by the Chodera lab using the LabMinds EasySolution/Revo to reproducibly prepare dialysis and experiment buffers, automated liquid handling to set up the experiments, and the automated calorimeter (Auto iTC-200) at the Rockefeller HTSRC across the street to execute them. From these experiments, we can obtain direct evidence of protonation state effects; in simple cases of a single relevant protonation state, we can extract both effective pK_a and affinity for each protonation state, but even in complex cases, experiment can be compared against simulated ITC data computed from the alchemical free energy calculation.

Selection criteria: For kinase:inhibitor complexes identified in **Aim 2B** as having proton shuttling events between complex and solvent upon binding (a net uptake/loss of protons), for which the inhibitor is sufficiently soluble in buffer to permit useful ITC experiments to be performed (≥250 μM), we will use this technique to experimentally determine whether proton shuttling events can be observed, and to experimentally validate the predicted difference in affinity between protonated and deprotonated species.

Potential pitfalls and alternative approaches: Protein stability can be a challenge in ITC experiments. We will perform Thermofluor melts over a range of buffer systems at pH 7.4 using the Chodera lab automation system [82] to determine buffers with differing ionization enthalpies in which the kinases are maximally stable.

Aim 3C. Use direct fluorescence binding affinity assays to examine binding events sensitive to pH

Tumor environments are often acidic due to increased aerobic and anaerobic glycolysis [84]. While intracellular pH is more tightly regulated than the extracellular pH, significant shifts in intracellular pH have been observed to range from 6.4–7.4 [84], which can manipulate the energetics of protonation states by up to 2.3 k_BT. For kinase:inhibitor complexes that show significant pH-dependence in computed binding affinity, we will measure pH-dependent binding affinities using a fluorescence-based affinity assay originally developed by collaborator Nicholas Levinson (University of Minnesota), who observed that inhibitors with certain scaffolds undergo a large increase in fluorescence upon binding to Src and Abl kinases [85]. The Chodera lab will use an automated plate-format fluorescence assay to measure the pH-dependent binding affinity of ATP-competitive inhibitors in the absence of ATP, using either their intrinsic fluorescence or competition with a fluorescent probe inhibitor (**Figure 4**).
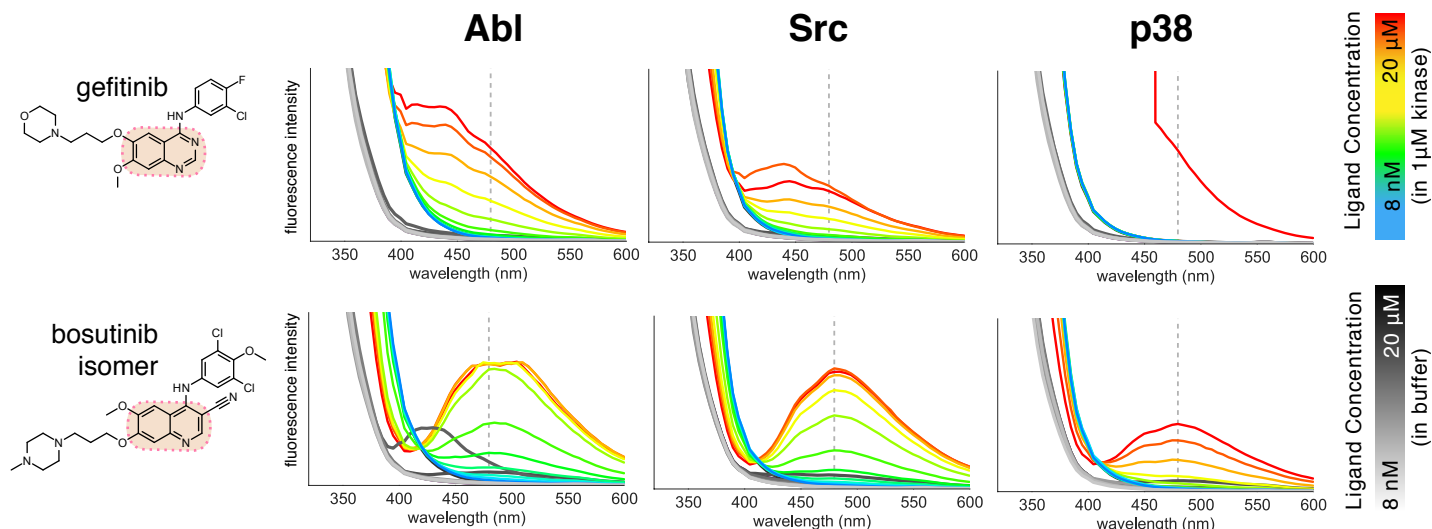
Figure 4. **Automated high-accuracy fluorescence assay for measuring direct binding of fluorescent kinase inhibitor probes.** By using ATP-competitive kinase inhibitor probes that greatly increase fluorescence upon binding, we can rapidly and directly measure the affinity of the probes over a very large dynamic range (nM–mM) or measure other inhibitor affinities by competition assays. While this phenomenon was initially described using a cuvette-based assay [85], we have developed it into an automated high-throughput assay. *Left:* Gefitinib (quinazoline scaffold) and an isomer of bosutinib (quinoline scaffold), just two of a large number of fluorescent probes that can be used for affinity measurement in this assay. *Right:* Fluorescence emission spectra (excitation at 280 nm) for different concentrations of probe compound shows large increase in fluorescence upon binding to kinase (1 μM in 100 μL volume in 96-well plates). The dashed vertical line (480 nm) indicates emission wavelength used for high-throughput form of this assay, which allows even weak (mM) binding to be quantified. Numerous FDA-approved quinazoline, quinoline, and pyrazolopyrimidine kinase inhibitors are usable as probes.

Selection criteria: To assess the potential for significant changes in inhibitor effectiveness between unperturbed ($\sim$7.4) and acidified ($\sim$6.4) intracellular environments—and hence differences in kinase inhibitor efficacy between normal and tumor cells—we will use the tools developed in **Aim 2** to compute the effect of acidified intracellular pH on binding affinities. This will first be estimated from the binding free energy calculations performed in **Aim 2B** without additional simulation by computing the derivative of the computed binding affinity with respect to pH using MBAR (**Equation 2**) [6]. Systems that show significant pH-dependence will then be subjected to the same pH-dependent alchemical free energy calculation described in **Aim 3A** to determine the pH-dependent affinity. If the affinity change is sufficiently large, the kinase:inhibitor complex will be eligible for this class of experiments.

Potential pitfalls and alternative approaches: If the fluorescent scaffold of the inhibitor changes protonation state during the pH titration, an inhibitor that shows a strong fluorescence change upon binding may cease to do so above or below a critical pH, limiting the useful range of this assay. The pH-dependent solubility of kinase inhibitors may additionally limit range of affinities measurable by this assay.

Aim 3D. Evaluate whether clinically-characterized mutations might induce resistance by modulating pK_as.

In systems where protonation state effects are significant, the introduction or elimination of charged residues in the vicinity of the binding site can cause electrostatically-induced shifts in pK_a that indirectly modulate inhibitor binding affinities, potentially manifesting a new mechanism of resistance. Such mutations would alter the electrostatic environments of key titratable groups directly or stabilize alternative conformations with shifted pK_as (as has been observed for site-directed mutations [35]). Engineering such mutations into the kinase and utilizing the fluorescence inhibitor binding assay from **Aim 3C** for both wild-type and mutant kinases will reveal whether these mutations are capable of causing measurable alternations to inhibitor binding free energies.
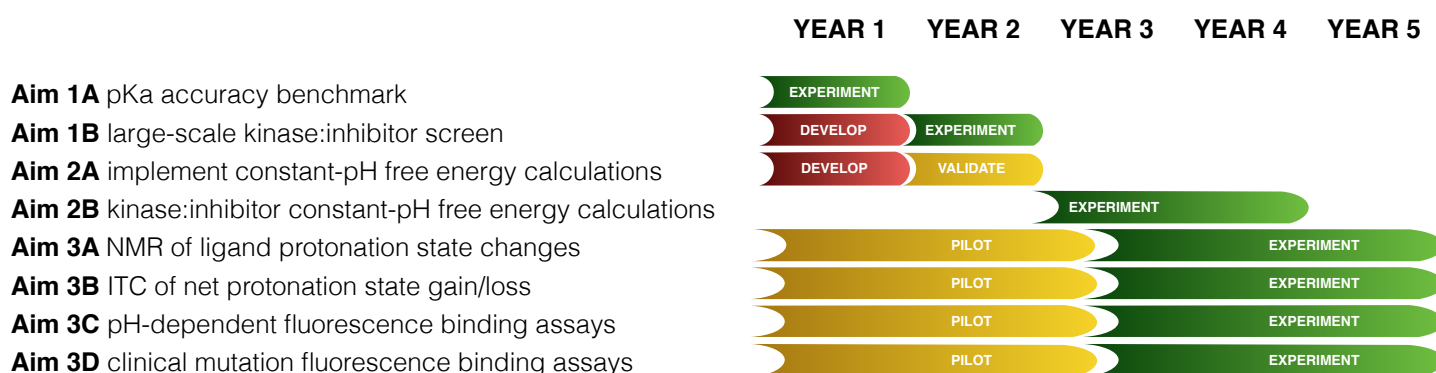
Selection criteria: We will again use the alchemical free energy calculation methodology developed in **Aim 2A** to computationally evaluate whether clinically-observed mutations (identified in the MSKCC cBioPortal) may cause measurable changes to inhibitor binding affinity. We restrict ourselves to kinases identified in **Aim 2B** to have significant protonation state effects and for which facile bacterial expression of the kinase permits site-directed mutagenesis, and for which direct changes in inhibitor fluorescence can be monitored due to the presence of a fluorescent scaffold (**Figure 4**). The Seeliger and Chodera labs will work together to engineer and express the mutants, and the Chodera lab will perform the fluorescence assays at pH 7.4 as in **Aim 3C**.

Potential pitfalls and alternative approaches: Should computational predictions differ wildly from the experimental measurements, this likely implicates other effects—such as multiple conformational states with slow conversion times—necessitating we turn to other computational techniques such as those developed by the Chodera laboratory and collaborators to study large-scale conformational changes in proteins [69].

OUTLOOK

This proposal seeks to both significantly advance our understanding of protonation state effects in the selective recognition of kinases by targeted inhibitors and develop the technology necessary to accurately treat these effects, thereby removing an enormous barrier to the quantiative prediction of small molecule binding affinities for rational design. By first broadly cataloging kinase:inhibitor complexes where protonation state effects may be prominent and then subsequently studying these systems in detail with a powerful new computational tool for treating protonation states dynamically in rigorous binding free energy calculations, we will be able to computationally dissect the types and magnitudes of protonation state effects. Subsequent experimental work in which pH is modulated or site-directed mutants are introduced will serve to confirm computational predictions and further elucidate implications of these effects. An increased understanding of these complex—but likely critical—effects will provide useful starting points for improved methodologies for discovering new selective kinase inhibitors for a variety of other kinase targets implicated in cancer. While this proposal only lays the groundwork for treating dynamic protonation states for the quantitative modeling of selective kinase inhibitors, this research will open the door to numerous other disease-specific routes of investigation in which protonation state effects could play key a critical role in disease mechanisms and effective kinase inhibition.

TIMELINE

	YEAR 1	YEAR 2	YEAR 3	YEAR 4	YEAR 5
Aim 1A pKa accuracy benchmark	EXPERIMENT				
Aim 1B large-scale kinase:inhibitor screen	DEVELOP	EXPERIMENT			
Aim 2A implement constant-pH free energy calculations	DEVELOP	VALIDATE			
Aim 2B kinase:inhibitor constant-pH free energy calculations			EXPERIMENT		
Aim 3A NMR of ligand protonation state changes	PILOT		EXPERIMENT		
Aim 3B ITC of net protonation state gain/loss	PILOT		EXPERIMENT		
Aim 3C pH-dependent fluorescence binding assays	PILOT		EXPERIMENT		
Aim 3D clinical mutation fluorescence binding assays	PILOT		EXPERIMENT		

COLLABORATION MANAGEMENT PLAN The Chodera, Gunner, and Seeliger groups are all located in close proximity, have worked together over the past three years, and already have an excellent collaborative relationship and rapport. In particular, Chodera and Gunner have worked together to develop new methodologies for constant-pH simulation, while Chodera and Seeliger have worked to develop new bacterially-expressing kinase constructs and a novel fluorescence assay for measuring kinase inhibitor binding affinities. To facilitate the close collaboration this project requires, we will utilize modern communication technologies such as Slack—a communication platform utilized by the JPL Mars Curiosity Rover and IceCube Neutrino Observatory teams—ensuring barrierless communication among members and PIs of all three groups. We will make use of shared public GitHub repositories to coordinate the development of simulation and analysis code and store simulation and experimental data. Frequent in-person visits between CCNY and MSKCC—the sites at which most software development will take place—are expected for the students heading software development. All project participants will receive accounts on the MSKCC High Performance Computing (HPC) resource, which provides a central location to carry out the computation for this project and share large datasets. We will hold monthly virtual meetings for subgroups working on the different aims and quarterly virtual meeting of all PIs and personnel, and semiannual physical meetings at MSKCC to review progress and plans in depth.

The PI (Chodera) will be responsible for the oversight and coordination of the overall project. The Gunner and Chodera labs will work closely on the identification of candidate kinase:inhibitor systems (**Aim 1**), with Co-I Gunner directing work related to the MCCE code produced in her laboratory. The Gunner and Chodera labs will also work closely on developing hybrid Monte Carlo / molecular dynamics algorithms for molecular dynamics simulation and alchemical free energy calculations and applying them to kinase:inhibitor systems (**Aim 2**), with Chodera directing this work. The Seeliger and Chodera labs will work closely on the experimental aspects of this project (**Aim 3**), with Co-I Seeliger directing experimental work in his laboratory (kinase expression, NMR) and Chodera directing experimental work in his laboratory (automated ITC, screening of mutants, fluorescence binding assays).

Bibliography and References Cited

[1] Mobley, D. L., Chodera, J. D., and Dill, K. A.: On the use of orientational restraints and symmetry corrections in alchemical free energy calculations. J. Chem. Phys. 125: 084902, 2006.

[2] Mobley, D. L., Chodera, J. D., and Dill, K. A.: Confine and release: Obtaining correct free energies in the presence of protein conformational change. J. Chem. Theor. Comput. 3: 1231–1235, 2007.

[3] Shirts, M. R., Mobley, D. L., Chodera, J. D., and Pande, V. S.: Accurate and efficient corrections for missing dispersion interactions in molecular simulations. J. Phys. Chem. B. 111: 13052–13063, 2007.

[4] Chodera, J. D., Singhal, N., Pande, V. S., Dill, K. A., and Swope, W. C.: Automatic discovery of metastable states for the construction of Markov models of macromolecular conformational dynamics. J. Chem. Phys. 126: 155101, 2007.

[5] Michael R. Shirts, D. L. M. and Chodera, J. D.: Alchemical free energy calculations: Ready for prime time? Annu. Rep. Comput. Chem. 3: 41–59, 2007.

[6] Shirts, M. R. and Chodera, J. D.: Statistically optimal analysis of samples from multiple equilibrium states. J. Chem. Phys. 129: 124105, 2008.

[7] Nilmeier, J. P., Crooks, G. E., Minh, D. D. L., and Chodera, J. D.: Nonequilibrium candidate Monte Carlo: A new tool for equilibrium simulation. Proc. Natl. Acad. Sci. USA. 108: E1009, 2011.

[8] Chodera, J. D. and Shirts, M. R.: Replica exchange and expanded ensemble simulations as gibbs sampling: Simple improvements for enhanced mixing. J. Chem. Phys. 135: 194110, 2011.

[9] Chodera, J. D., Mobley, D. L., Shirts, M. R., Dixon, R. W., Branson, K. M., and Pande, V. S.: Free energy methods in drug discovery and design: Progress and challenges. Curr. Opin. Struct. Biol. 21: 150, 2011.

[10] Wang, K., Chodera, J. D., Yang, Y., and Shirts, M. R.: Identifying ligand binding sites and poses using GPU-accelerated Hamiltonian replica exchange molecular dynamics. J. Comput. Aided Mol. Des. 27: 989, 2013.

[11] Alexov, E. and Gunner, M. R.: Incorporating protein conformational flexibility into pH-titration calculations: Results on T4 lysozyme. Biophys. J. 74: 2075–2093, 1997.

[12] Alexov, E. and Gunner, M. R.: A pragmatic approach to structure based calculation of coupled proton and electron transfer in proteins. Biochim. Biophys. Acta. 1458: 63–87, 2000.

[13] Georgescu, R. E., Alexov, E. G., and Gunner, M. R.: Combining conformational flexibility and continuum electrostatics for calculating pK_as in proteins. Biophys. J. 83: 1731–1748, 2002.

[14] Song, Y., Mao, J., and Gunner, M. R.: MCCE2: Improving protein pK_a calculations with extensive side chain rotamer sampling. J. Comput. Chem. 30: 2231–2247, 2009.

[15] Song, Y. and Gunner, M. R.: Using multiconformation continuum electrostatics to compare chloride binding motifs in α-amylase, human serum albumin, and Omp32. J. Mol. Biol. 387: 840—846, 2009.

[16] Zheng, Z., Dutton, P. L., and Gunner, M. R.: The measured and calculated affinity of methyl and methoxy substituted benzoquinones for the QA site of bacterial reaction centers. Proteins. 78: 2638–2654, 2010.

[17] Warncke, K., Gunner, M. R., Braun, B. S., Gu, L., Yu, C., Bruce, J. M., and Dutton, P. L.: Influence of hydrocarbon tail structure on quinone binding and electron-transfer performance at the QA and QB sites of the photosynthetic reaction center protein. Biochem. 33: 7830–7841, 1994.

[18] Zhang, X. and Gunner, M. R.: Affinity and activity of non-native quinones at the Q(B) site of bacterial photosynthetic reaction centers. Photosynth. Res. 120: 181–196, 2013.

[19] Nielsen, J. E., Gunner, M. R., and Garcia-Moreno, E. B.: The pK_a cooperative: a collaborative effort to advance structure-based calculations of pka values and electrostatic effects in proteins. Proteins. 79: 3249–3259, 2011.

[20] Stanton, C. L. and Houk, K. N.: Benchmarking pK_a prediction methods for residues in proteins. J. Chem. Theory Comput. 4: 951–966, 2008.

[21] Szakács, Z., Béni, S., Varga, Z., Örfi, L., Kéri, G., and Noszál, B.: Acid-base profiling of imatinib (Gleevec) and its fragments. J. Med. Chem. 48: 249, 2005.

[22] American Cancer Society, . Cancer Facts & Figures 2014, 2014.

[23] Zhang, J., Yang, P. L., and Gray, N. S.: Targeting cancer with small molecule kinase inhibitors. Nat. Rev. Cancer. 9: 28–39, 2009.

[24] Deininger, M., Buchdunger, E., and Druker, B. J.: The development of imatinib as a therapeutic agent for chronic myeloid leukemia. Blood. 105: 2640–2653, 2005.

[25] Stegmeier, F., Warmuth, M., Sellers, W. R., and Dorsch, M.: Targeted cancer therapies in the twenty-first century: Lessons from imatinib. Clin. Pharm. & Therap. 87: 543–552, 2010.

[26] Longley, D. B. and Johnston, P. G.: Molecular mechanisms of drug resistance. J. Pathol. 205: 275–292, 2005.

[27] The story of Gleevec. http://www.innovation.org/index.cfm/StoriesofInnovation/InnovatorStories/

`The_Story_of_Gleevec`. Accessed: 2014-05-03.

[28] Shirts, M. R., Mobley, D. L., and Brown, S. P. Free energy calculations in structure-based drug design. In Structure Based Drug Design. Cambridge University Press, 2009.

[29] Paul, S. M., Mytelka, D. S., Dunwiddie, C. T., Persinger, C. C., Munos, B. H., Lindborg, S. R., and Schacht, A. L.: How to improve R&D productivity: The pharmaceutical industry's grand challenge. Nature Rev. Drug Discovery. 9: 203, 2010.

[30] Schneider, G.: Virtual screening: an endless staircase? Nature Rev. Drug Discovery. 9: 273–276, 2010.

[31] ten Brink, T. and Exner, T. E.: pk_a based protonation states and microspecies for protein-ligand docking. J. Comput. Aided Mol. Des. 24: 935–942, 2010.

[32] Onufriev, A. V. and Alexov, E.: Protonation and pK changes in protein-ligand binding. Quart. Rev. Biophys. 46: 181–209, 2013.

[33] Seeliger, M. A., Nagar, B., Frank, F., Cao, X., Henderson, M. N., and Kuriyan, J.: c-Src binds to the cancer drug imatinib with an inactive Abl/c-Kit conformation and a distributed thermodynamic penalty. Structure. 15: 299–311, 2007.

[34] Lin, Y.-L., Meng, Y., Jiang, W., and Roux, B.: Explaining why Gleevec is a specific and potent inhibitor of Abl kinase. Proc. Natl. Acad. Sci. USA. 110: 1664–1669, 2013.

[35] Shan, Y., Seeliger, M. A., Eastwood, M. P., Frank, F., Xu, H., Jensen, M. O., Dror, R. O., Kuriyan, J., and Shaw, D. E.: A conserved protonation-dependent switch controls drug binding in Abl kinase. Proc. Natl. Acad. Sci. USA. 106: 139–144, 2009.

[36] Aleksandrov, A. and Simonson, T.: A molecular mechanics model for imatinib and imatinib:kinase binding. J. Comput. Chem. 31: 1550–1560, 2010.

[37] Aleksandrov, A. and Simonson, T.: Molecular dynamics simulations show that conformational selection governs the binding preferences of imatinib for several tyrosine kinases. J. Biol. Chem. 285: 13807–13815, 2010.

[38] Park, M.-S., Gao, C., and Stern, H. A.: Estimating binding affinities by docking/scoring methods using variable protonation states. Proteins: Structure, Function, and Bioinformatics. 79: 304–314, 2010.

[39] Mongan, J., Case, D. A., and McCammon, J. A.: Constant pH molecular dynamics in a generalized born implicit solvent. J. Comput. Chem. 25: 2038–2048, 2004.

[40] Khandogin, J. and Brooks, C. L.: Constant pH molecular dynamics with proton tautomerism. Biophys. J. 89: 141–157, 2005.

[41] Stern, H. A.: Molecular simulation with variable protonation states at constant pH. J. Chem. Phys. 126: 164112, 2007.

[42] Donnini, S., Tegeler, F., Groenhof, G., and Grubmüller, H.: Constant pH molecular dynamics in explicit solvent with λ-dynamics. J. Chem. Theor. Comput. 7: 1962–1978, 2011.

[43] Goh, G. B., Hulbert, B. S., Zhou, H., and Brooks III, C. L.: Constant pH molecular dynamics of proteins in explicit solvent with proton tautomerism. Proteins. 82: 1319–1331, 2014.

[44] Chen, Y. and Roux, B.: Constant-pH hybrid nonequilibrium molecular dyanamics-Monte Carlo simulation method. J. Chem. Theor. Comput. 11: 3919–3931, 2015.

[45] Lee, J., Miller, B. T., and Brooks, B. R.: Computational scheme for pH-dependent binding free energy calculation with explicit solvent. Protein Sci. 25: 231–243, 2015.

[46] Wang, L., Wu, Y., Deng, Y., Kim, B., Pierce, L., Krilov, G., Lupyan, D., Robinson, S., Dahlgren, M. K., Greenwood, J., Romero, D. L., Masse, C., Knight, J. L., Steinbrecher, T., Beuming, T., Damm, W., Harder, E., Sherman, W., Brewer, M., Wester, R., Murcko, M., Frye, L., Farid, R., Lin, T., Mobley, D. L., Jorgensen, W. L., Berne, B. J., Friesner, R. A., and Abel, R.: Accurate and reliable prediction of relative ligand binding potency in prospective drug discovery by way of a modern free-energy calculation protocol and force field. J. Am. Chem. Soc. 137: 2695–2703, 2015.

[47] Dullweber, F., Stubbs, M. T., Musil, Đ., Stürzebecher, J., and Klebe, G.: Factorising ligand affinity: A combined thermodynamic and crystallographic study of trypsin and thrombin inhibition. J. Mol. Biol. 313: 593–614, 2001.

[48] Stubbs, M. T., Reyda, S., Dullweber, F., Möller, M., Klebe, G., Dorsch, D., Mederski, W. W. K. R., and Wurziger, H.: pH-dependent binding modes observed in trypsin crystals: Lessons for structure-based drug design. ChemBioChem. 02-03: 246, 2002.

[49] Czdrowski, P., Sotriffer, C. A., and Klebe, G.: Protonation changes upon ligand binding to trypsin and thrombin: Structural interpretation based on pk_a calculations and ITC experiments. J. Mol. Biol. 367: 1347–1356, 2007.

[50] Steuber, H., Czdrowski, P., Sotriffer, C. A., and Klebe, G.: Tracing changes in protonation: A prerequisite to factorize thermodynamic data of inhibitor binding to aldose reductase. J. Mol. Biol. 373: 1305–1320, 2007.

[51] Dashi, D. S., Meng, Y., and Roitberg, A. E.: pH-replica exchange molecular dynamics in proteins using a

discrete protonation method. J. Phys. Chem. B. 116: 8805–8811, 2012.

[52] Khandogin, J. and Brooks III, C. L.: Constant pH molecular dynamics with proton tautomerism. Biophys. J. 89: 141–157, 2005.

[53] Tembe, B. L. and McCammon, J. A.: Ligand-receptor interactions. Computers & Chemistry. 8: 281–283, 1984.

[54] Michel, J. and Essex, J.: Hit identification and binding mode predictions by rigorous free energy simulations. J. Med. Chem. 51: 6654–6664, 2008.

[55] Michel, J. and Essex, J. W.: Prediction of protein–ligand binding affinity by free energy simulations: assumptions, pitfalls and expectations. J. Comput. Aided Mol. Des. 24: 639–658, 2010.

[56] Gallicchio, E. and Levy, R. M.: Advances in all atom sampling methods for modeling protein–ligand binding affinities. Curr. Opin. Struct. Biol. 21: 161–166, 2011.

[57] Yank: A gpu-accelerated python framework for exploring algorithms for alchemical free energy calculations. http://getyank.org.

[58] Eastman, P., Friedrichs, M., Chodera, J. D., Radmer, R., Bruns, C., Ku, J., Beauchamp, K., Lane, T. J., Wang, L.-P., Shukla, D., Tye, T., Houston, M., Stitch, T., and Klein, C.: OpenMM 4: A reusable, extensible, hardware independent library for high performance molecular simulation. J. Chem. Theor. Comput. 9: 461, 2012.

[59] Openmm: A high performance toolkit for molecular simulation. http://openmm.org.

[60] Minh, D. D. L. and Chodera, J. D.: Estimating equilibrium ensemble averages using multiple time slices from driven nonequilibrium processes: theory and application to free energies, moments, and thermodynamic length in single-molecule pulling experiments. J. Chem. Phys. 134: 024111, 2011.

[61] Minh, D. D. L. and Chodera, J. D.: Optimal estimators and asymptotic variances for nonequilibrium path-ensemble averages. J. Chem. Phys. 131(13): 134110, 2009.

[62] Seeliger, M. A., Young, M., Henderson, M. N., Pellicena, P., King, D. S., Falick, A. M., and Kuriyan, J.: High yield bacterial expression of active c-Abl and c-Src tyrosine kinases. Protein Sci. 14: 3135–3139, 2005.

[63] Shelley, J. C., Cholleti, A., Frye, L. L., Greenwood, J. R., Timlin, M. R., and Uchiyama, M.: Epik: A software program for pKa prediction and protonation state generation for drug-like molecules. J. Comput. Aided Mol. Des. 21: 681–691, 2007.

[64] Milletti, F., Storchi, L., and Cruciani, G.: Predicting protein pk_a by environment similarity. Proteins. 50: 484–495, 2009.

[65] Jakalian, A., Bush, B. L., Jack, D. B., and Bayly, C. I.: Fast, efficient generation of high-quality atomic charges. AM1-BCC model: I. method. J. Comput. Chem. 21(2): 132–146, 2000.

[66] Jakalian, A., Jack, D. B., and Bayly, C. I.: Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. parameterization and validation. J. Comput. Chem. 23(16): 1623–1641, 2002.

[67] Qian, B., Raman, S., Das, R., Bradley, P., McCoy, A. J., Read, R. J., and Baker, D.: High-resolution structure prediction and the crystallographic phase problem. Nature. 450(7167): 259–264, November 2007.

[68] Wang, C., Bradley, P., and Baker, D.: Protein–Protein Docking with Backbone Flexibility. Journal of Molecular Biology. 373(2): 503–519, October 2007.

[69] Prinz, J.-H., Wu, H., Sarich, M., Keller, B., Fischbach, M., Held, M., Chodera, J. D., Schütte, C., and Noé, F.: Markov models of molecular kinetics: Generation and validation. J. Chem. Phys. 134: 174105, 2011.

[70] Crooks, G. E.: Measuring thermodynamic length. Phys. Rev. Lett. 99(10): 100602, 2007.

[71] Shenfeld, D. K., Xu, H., Eastwood, M. P., Dror, R. O., and Shaw, D. E.: Minimizing thermodynamic length to select intermediate states for free-energy calculations and replica-exchange simulations. Phys. Rev. E. 80(4): 046705, 2009.

[72] Swope, W. C., Andersen, H. C., Berens, P. H., and Wilson, K. R.: A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. J. Chem. Phys. 76(1): 637–649, 1982.

[73] Lyubartsev, A. P., Martsinovski, A. A., Shevkunov, S. V., and Vorontsov-Velyaminov, P. N.: New approach to Monte Carlo calculation of the free energy: Method of expanded ensembles. J. Chem. Phys. 96: 1776–1783, 1992.

[74] Tan, Z.: Optimally adjusted mixture sampling and locally weighted histogram analysis. J. Comput. Graph. Stat. in press.

[75] Bain, J., Plater, L., Elliott, M., Shpiro, N., Hastie, C. J., McLauchlan, H., Klevernic, I., Arthur, J. S. C., Aless, D. R., and Cohen, P.: The selectivity of protein kinase inhibitors: A further update. Biochem. J. 408: 297–315, 2007.

[76] Fabian, M. A., Biggs, W. H., Treiber, D. K., Atteridge, C. E., Azimioara, M. D., Benedetti, M. G., Carter, T. A., Ciceri, P., Edeen, P. T., Floyd, M., Ford, J. M., Galvin, M., Gerlach, J. L., Grotzfeld, R. M., Herrgard, S., Insko, D. E., Insko, M. A., Lai, A. G., Lélias, J.-M., Mehta, S. A., Milanov, Z. V., Velasco, A. M., Wodicka, L. M.,

Patel, H. K., Zarrinkar, P. P., and Lockhart, D. J.: A small molecule–kinase interaction map for clinical kinase inhibitors. Nature Biotech. 23: 329–336, 2005.

[77] Karaman, M. W., Herrgard, S., Treiber, D. K., Gallant, P., Atteridge, C. E., Campbell, B. T., Chan, K. W., Ciceri, P., Davis, M. I., Edeen, P. T., Faraoni, R., Floyd, M., Hunt, J. P., Lockhart, D. J., Milanov, Z. V., Morrison, M. J., Pallares, G., Patel, H. K., Pritchard, S., Wodicka, L. M., and Zarrinkar, P. P.: A quantitative analysis of kinase inhibitor selectivity. Nature Biotech. 26: 127–132, 2008.

[78] Parton, D. L., Grinaway, P. B., Hanson, S. M., Beauchamp, K. A., and Chodera, J. D.: Ensembler: Enabling high-throughput molecular simulations at the superfamily scale. PLoS Computational Biology. in press.

[79] We will start with AMBER family protein forcefields ff99sb-ildn-phi [86] and ff99sb-ildn-nmr [87], which have been shown to give excellent agreement with NMR data [88] and the generalized AMBER small molecule forcefield (GAFF) [89]. We will initially use the TIP3P [90] and TIP4P-Ew [91] explicit solvent models.

[80] Parton, D. L., Hanson, S. M., Rodríguez-Laureano, L., Albanese, S. K., Jeans, C., Gradia, S., Seeliger, M. A., and Chodera, J. D.: A panel of recombinant human kinase domain constructs for automated bacterial expression. bioRxiv. http://dx.doi.org/10.1101/038711 pp, 2016.

[81] Meyer, B. and Peters, T.: Nmr spectroscopy techniques for screening and identifying ligand binding to protein receptors. Angew. Chem. Int. Ed. Engl. 42: 864–890, 2003.

[82] Lavinder, J. L., Hari, S. B., Sullivan, B. J., and Magliery, T. J.: High-throughput thermal scanning: a general, rapid dye-binding thermal shift screen for protein engineering. J. Am. Chem. Soc. 131: 3794–3795, 2009.

[83] Baker, B. M. and Murphy, K. P.: Evaluation of linked protonation effects in protein binding reactions using isothermal titration calorimetry. Biophys. J. 71: 2049–2055, 1996.

[84] Song, C. W., Griffin, R., and Park, H. J. Cancer Drug Discovery and Development, chapter Influence of Tumor pH on Therapeutic Response, pp 21–42. Humana Press, Totowa, New Jersey, 2006.

[85] Levinson, N. M. and Boxer, S. G.: Structural and spectroscopic analysis of the kinase inhibitor bosutinib and an isomer of bosutinib binding to the Abl tyrosine kinase domain. PLoS One. 7: e29828, 2012.

[86] Nerenberg, P. and Head-Gordon, T.: Optimizing protein-solvent force fields to reproduce intrinsic conformational preferences of model peptides. J. Chem. Theor. Comput. 7: 1220–1230, 2011.

[87] Li, D.-W. and Brüschweiler, R.: NMR-based protein potentials. Angew. Chemie. 38: 6930–6932, 2010.

[88] Beauchamp, K. A., Lin, Y.-S., Das, R., and Pande, V. S.: Are protein force fields getting better? a systematic benchmark on 524 diverse NMR measurements. J. Chem. Theor. Comput. 8: 1409–1414, 2012.

[89] Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A.: Development and testing of a general AMBER force field. J. Comput. Chem. 25: 1157–1174, 2004.

[90] Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L.: Comparison of simple potential functions for simulating liquid water. J. Chem. Phys. 79: 926, 1983.

[91] Horn, H. W., Swope, W. C., Pitera, J. W., Madura, J. D., Dick, T. J., Hura, G. L., and Head-Gordon, T.: Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. J. Chem. Phys. 120: 9665, 2004.